GLAMpoints: Greedily Learned Accurate Match points
SUPPLEMENTARY MATERIAL

Prune Truong[1,2*] Stefanos Apostolopoulos[1] Agata Mosinska[1] Samuel Stucky[1] Carlos Ciller[1]
Sandro De Zanet[1]

[1]RetinAI Medical AG, Switzerland [2]ETH Zurich, Switzerland
truongp@student.ethz.ch {stefanos, agata, samuel, carlos, sandro}@retinai.com

1. Supplementary details on the training method

1.1. Performance comparison between SIFT descriptor with/without rotation invariance

Greedily Learned Accurate Match Points (GLAMpoints) detector was trained and tested in association with Scale-Invariant Feature Transform (SIFT) descriptor rotation-dependent because SIFT descriptor without rotation invariance performs better than the rotation invariant version on fundus images. The details of the metrics evaluated on the pre-processed *slitlamp* dataset for both versions of SIFT descriptor are shown in Table 1.

Table 1: Metrics calculated over 206 pre-processed pairs of the *slitlamp* dataset. Best results of each category are indicated in bold.

	SIFT with rotation invariance	SIFT rotation-dependent
Success rate of Acceptable Registrations [%]	49.03	**50.49**
Success rate of Inaccurate Registrations [%]	50.49	47.57
Success rate of Failed Registrations [%]	0.49	1.94
M.score	0.0470	**0.056**
Coverage Fraction	0.1348	**0.15**
AUC	0.1274	**0.143**

1.2. Method for homography generation

Let B be the set of base images of size $H \times W$, used for training. At every step i, an image pair I_i, I_i' is generated from an original image B_i by applying two separate, randomly sampled homography transforms g_i, g_i'. Each of those homographies is a composition of rotation, shearing, perspective, scaling and translation elements. The minimum and maximum values of the parameters are given in table 2.

Table 2: Parameters used for random homography generation during training.

Scaling		**Perspective**		**Translation**		**Shearing**		**Rotation**	
min scaling	0.7	min perspective parameter	0.000001	max horizontal displacement	100	min/max horizontal shearing	-0.2 / 0.2	max angle	25
max scaling	1.3	max perspective parameter	0.0008	max vertical displacement	100	min/max vertical shearing	-0.2 / 0.2		

*Work produced during an internship at RetinAI Medical AG

2. Details of results on fundus images

2.1. Details of *MEE* and *RMSE* per registration class on the retinal images dataset

Table 3 and 4 shows the mean and standard deviation of MEE and $RMSE$ for respectively the *FIRE* dataset and the $slitlamp$ dataset. In both cases, GLAMpoints (NMS10) presents the highest registration accuracy for inaccurate registrations and globally.

Table 3: Means and standard deviations of median errors (MEE) and RMSE in pixels for non-preprocessed images of the *FIRE* dataset. Acceptable registrations are defined as having ($MEE < 10$ and $MAE < 30$). Best results per category are indicated in bold.

	Inaccurate Registration		**Acceptable Registration**		**Global Non-Failed Registration**	
	MEE	*RMSE*	*MEE*	*RMSE*	*MEE*	*RMSE*
SIFT	66.44 ± 86.98	179.2 ± 412.88	3.79 ± 2.27	3.97 ± 2.3	27.22 ± 61.25	69.51 ± 266.37
KAZE	105.36 ± 118.94	314.0 ± 1184.76	4.69 ± 2.2	4.53 ± 2.28	72.97 ± 108.66	214.43 ± 986.38
SuperPoint	33.79 ± 69.65	48.97 ± 134.77	2.19 ± 2.12	2.2 ± 2.22	6.44 ± 27.78	8.48 ± 51.94
LIFT	24.39 ± 46.91	25.97 ± 43.93	2.4 ± 2.26	2.48 ± 2.54	4.7 ± 16.72	4.94 ± 16.09
LF-Net						
GLAMpoints (OURS)	**15.53 ± 7.89**	**16.42 ± 6.26**	**2.58 ± 2.36**	**2.74 ± 2.54**	**3.26 ± 4.1**	**3.46 ± 4.17**

Table 4: Means and standard deviations of median errors (MEE) and RMSE in pixels for the 206 images of the $slitlamp$ dataset. Acceptable registrations are defined as having ($MEE < 10$ and $MAE < 30$). Best results per category are indicated in bold.

a) Non pre-processed data

	Inaccurate Registration		**Acceptable Registration**		**Global Non-Failed Registration**	
	MEE	*RMSE*	*MEE*	*RMSE*	*MEE*	*RMSE*
SIFT	109.04 ± 132.13	368.56 ± 1766.18	5.15 ± 2.37	5.78 ± 2.45	81.89 ± 122.39	273.74 ± 1526.27
KAZE	139.12 ± 123.07	640.8 ± 2980.66	5.58 ± 2.66	5.72 ± 2.39	114.29 ± 122.6	522.74 ± 2700.71
SuperPoint	131.82 ± 123.28	231.08 ± 509.82	**3.82 ± 1.78**	**3.77 ± 1.71**	79.12 ± 113.62	137.48 ± 406.7
LIFT	114.25 ± 129.96	1335.03 ± 10820.78	3.94 ± 2.08	4.04 ± 2.04	52.14 ± 101.86	585.54 ± 7182.71
LF-NET	77.69 ± 112.34	92.97 ± 183.92	4.61 ± 2.28	4.62 ± 2.31	33.7 ± 79.41	39.79 ± 123.85
GLAMpoints (OURS)	**25.77 ± 38.32**	**33.15 ± 85.49**	4.61 ± 2.16	4.6 ± 2.26	**12.32 ± 25.32**	**15.0 ± 53.41**

b) Pre-processed data

	Inaccurate Registration		**Acceptable Registration**		**Global Non-Failed Registration**	
	MEE	*RMSE*	*MEE*	*RMSE*	*MEE*	*RMSE*
SIFT	65.2 ± 90.35	130.55 ± 273.75	4.92 ± 2.15	5.01 ± 2.25	34.17 ± 69.79	65.92 ± 200.74
KAZE	86.83 ± 117.22	870.24 ± 7016.58	4.33 ± 2.26	4.45 ± 2.43	50.26 ± 96.6	486.39 ± 5252.64
SuperPoint	117.53 ± 125.01	194.5 ± 312.9	4.21 ± 2.03	4.11 ± 2.05	67.43 ± 109.03	110.33 ± 252.12
LIFT	113.3 ± 134.58	1328.06 ± 8854.49	**4.15 ± 2.25**	**4.21 ± 2.36**	47.6 ± 100.34	531.18 ± 5623.92
LF-NET	75.34 ± 128.6	158.78 ± 473.55	4.41 ± 2.16	4.45 ± 2.23	30.58 ± 85.3	61.39 ± 297.12
GLAMpoints (OURS)	**30.13 ± 56.86**	**27.53 ± 42.41**	4.85 ± 2.44	4.85 ± 2.47	**12.83 ± 34.09**	**12.01 ± 26.13**

2.2. Supplementary examples of matching on the FIRE dataset

Matches between two pairs of the $FIRE$ dataset are shown in Figure 1 for GLAMpoints, SIFT, KAZE, SuperPoint, Learned Invariant Feature Transform (LIFT) and Local Feature Network (LF-NET).

3. Generalization of the model on natural images

Our method was tested on several natural image datasets, with following specifications:

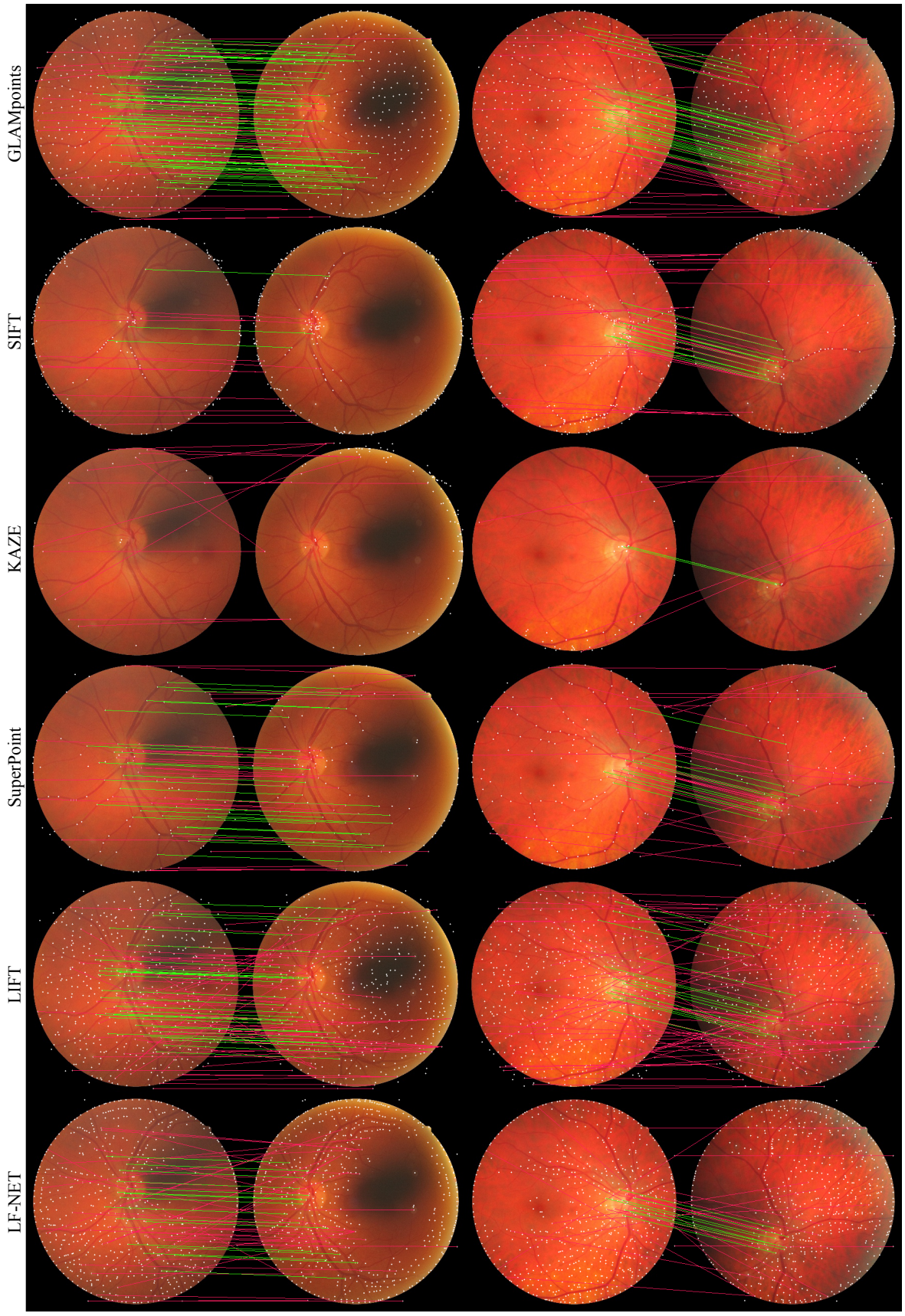

Figure 1: Matches on the $FIRE$ dataset. Detected points are in white, green lines are true positive matches while red ones are false positive.

1. $Oxford$ dataset: 8 sequences with 45 pairs in total. The dataset contains various imaging changes including viewpoint, rotation, blur, illumination, scale, JPEG compression changes. We evaluated on six of these sequences, excluding the ones showing rotation. Indeed, we trained our model associated with SIFT descriptor without rotation invariance. To be consistent, SIFT descriptor rotation-dependent was also used for testing.

2. $ViewPoint$ dataset: 5 sequences with 25 pairs in total. It exhibits large viewpoint changes and in-plane rotations up to 45 degrees.

3. EF dataset: 3 sequences with 17 pairs in total. The dataset exhibits drastic lighting changes as well as daytime changes and viewpoint changes.

4. $Webcam$ dataset: 6 sequences with 124 pairs in total. It shows seasonal changes as well as day time changes of scene taken from far away.

The metrics computed on those datasets are shown in Figure 2. We use the same thresholds as in the main paper to determine successful, inaccurate and failed registration. We used the LF-NET pretrained on outdoor data, since most images of those datasets are outdoor. It is worth mentioning the gap in performance between SIFT descriptor with or without rotation invariance on the EF and the $Viewpoints$ datasets. Those images exhibit large rotations and therefore a rotation invariant descriptor is necessary, which is not currently the case of our detector associated with SIFT. This explains why GLAMpoints performs poorly on those datasets. It is interesting to note that LF-NET scores extremely low in all metrics except for $repeatability$ on the *Viewpoints* dataset, because it finds only very few true positive matches compared to the number of detected keypoints and matches.

On the $Oxford$ dataset, GLAMpoints outperforms all others in terms of $M.score$, *coverage fraction* and AUC while scoring second in repeatability.

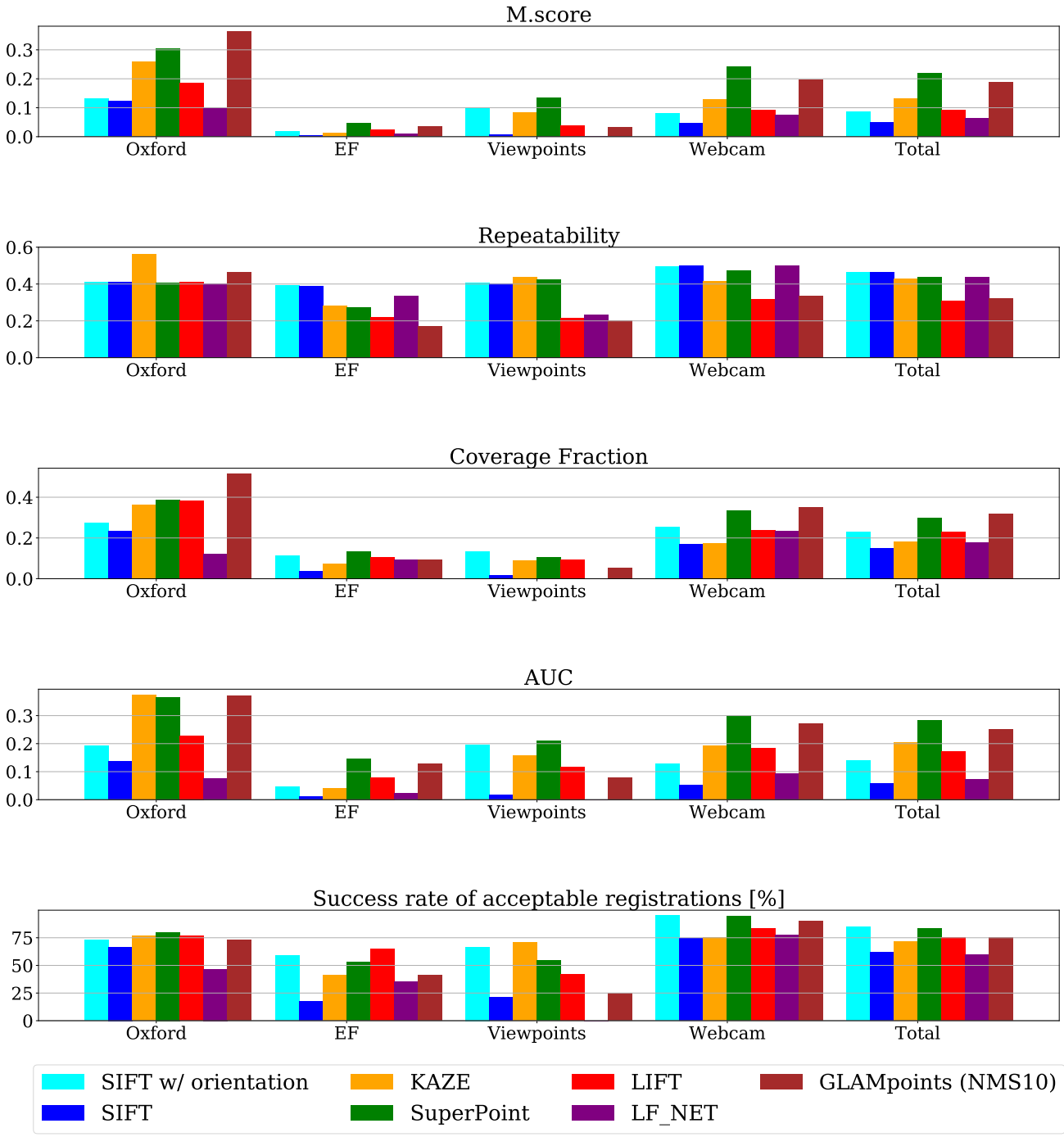

Figure 2: Summary of detector/descriptor performance metrics evaluated over 195 pairs of natural images.